



ACADEMIC
PRESS

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Journal of Memory and Language 49 (2003) 43–61

Journal of
Memory and
Language

www.elsevier.com/locate/jml

The effects of common ground and perspective on domains of referential interpretation[☆]

Joy E. Hanna,^{a,*} Michael K. Tanenhaus,^b and John C. Trueswell^c

^a *Department of Psychology, State University of New York at Stony Brook, Stony Brook, NY 11794, USA*

^b *Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY, USA*

^c *Department of Psychology, University of Pennsylvania, Philadelphia, PA, USA*

Received 30 May 2002; revision received 8 January 2003

Abstract

Addressees' eye movements were tracked as they followed instructions given by a confederate speaker hidden from view. Experiment 1 used objects in common ground (known to both participants) or privileged ground (known to the addressee). Although privileged objects interfered with reference to an identical object in common ground, addressees were always more likely to look at an object in common ground than privileged ground. Experiment 2 used definite and indefinite referring expressions with early or late points of disambiguation, depending on the uniqueness of the display objects. The speaker's and addressee's perspectives matched when the speaker was accurately informed about the display, and mismatched when the speaker was misinformed. When perspectives matched, addressees identified the target faster with early than with late disambiguation displays. When perspectives mismatched, addressees still identified the target quickly, showing an ability to use the speaker's perspective. These experiments demonstrate that although addressees cannot completely ignore information in privileged ground, common ground and perspective each have immediate effects on reference resolution.

© 2003 Elsevier Science (USA). All rights reserved.

Keywords: Reference resolution; Common ground; Perspective; Conversation; Communication; Language

In order to communicate successfully, conversational partners must coordinate their individual knowledge and actions with one another in order to create and interpret linguistic expressions. However, until recently models of the time course of language processing have tended to ignore the effects of being engaged in

conversation (cf. Clark, 1997). Thus relatively little is known about how conversational context affects the moment-by-moment processes central to real-time language production and comprehension. The current experiments monitored eye movements in referential communication tasks. They explored the time course

[☆] This material is based upon work supported by the National Institutes of Health under National Research Service Award 1F32MH1263901A1 and Grant No. HD-27206, and by the National Science Foundation under SBR Grant No. 95-11340, and Grant Nos. 0082602 and 9980013. We would like to thank Martin Pickering and two anonymous reviewers for their helpful comments. We are very grateful to Jared Novick, Ellen Hogan, and Craig Chambers for being the confederates, and to Jared Novick and Eman Moustafa for assistance in data collection and coding. Some of the research in this paper was reported at the Nineteenth Annual Conference of the Cognitive Science Society (August, 1997) in Palo Alto, CA, the Thirty-Eighth Annual Meeting of the Psychonomic Society (November, 1997) in Philadelphia, PA, the Eleventh Annual CUNY Conference on Human Sentence Processing (March, 1998) in Rutgers, NJ, and the Fourteenth Annual CUNY Conference on Human Sentence Processing (March, 2001) in Philadelphia, PA.

* Corresponding author.

E-mail address: jhanna@sunysb.edu (J.E. Hanna).

with which addressees can use information arising from the context of interacting with a speaker during reference resolution, focusing primarily on definite reference.

Definite reference is an important class of referring expression. Definite noun phrases, for example, noun phrases beginning with definite determiners, such as *the*, *this*, and *that*, and definite pronouns, such as *he*, *she*, and *it*, carry a presupposition of a referent that is uniquely identifiable with respect to a particular context (cf. Searle, 1969; but also see Epstein, 1998, for the use of definiteness to signal prominence). In fact, the interpretation of definite reference must take place with respect to a particular context; otherwise there would be a practically unlimited number of available interpretations (e.g., the rather large set of all possible referents of *the experiment* that exist in the world). Since uniquely identifying an entity denoted by a definite reference happens with respect to a particular and limited context (e.g., Barwise, 1989; Barwise & Perry, 1983), two questions arise: (1) what determines or constitutes the appropriate context, or *domain of interpretation* and (2) what is the time course with which this context is coordinated with other linguistic systems to have appropriately restrictive effects on language processing?

One proposal, perhaps most notably made by Clark and his colleagues, is that a particular kind of context, the *common ground* or mutual knowledge, beliefs, and assumptions among conversational participants (see Clark, 1992 and Smith, 1982 for discussion about defining common ground), has a primary role in defining the domain of interpretation (e.g., Clark, 1992, 1996; Clark & Brennan, 1989; Clark & Marshall, 1981; Clark & Shaefer, 1987; Clark & Wilkes-Gibbs, 1986). Common ground includes information that comes from the bases of community membership, physical co-presence, and linguistic co-presence. For example, conversational participants would be able to infer that they share various types of knowledge on the basis of both being in a particular city, or by looking at a particular object at the same time, or by maintaining a record of what has been discussed. The process of collaboratively establishing and maintaining common ground plays a crucial role in determining many aspects of reference, from whether it will be definite or indefinite (Clark & Haviland, 1977), to the precise linguistic form it will take (e.g., Brennan, 1990; Brennan & Clark, 1996; Clark & Clark, 1979; Clark & Wilkes-Gibbs, 1986; Lockridge & Brennan, 2002).

As a consequence of these results, Clark and colleagues have come to view language as a type of joint action that is performed by people who cooperate with each other in order to achieve particular goals (Clark, 1996). Therefore, the mutual knowledge between conversational participants becomes the playing field, or the relevant context, against which this cooperation can take place; that is, interlocutors can cooperate and

communicate successfully only if they monitor what is mutually known about a situation and use that knowledge effectively.

A logical consequence of the joint action perspective is that one of the primary roles of common ground is to act as the domain of interpretation for reference. Indeed, Clark poses the possibility that if the language processing system is optimal, the domain of interpretation for definite referring expressions will be restricted to entities in common ground, and will not include information that is not mutually known (Clark, 1992). However, while common ground ultimately does determine the appropriate context for interpretation, the timing of its effects remains an open empirical question. This is in part due to the methods used to investigate conversational interaction. Typically, naïve subjects are placed in referential communication tasks where they converse with one another while performing a matching task with complex figures, such as Tangrams (for the original Tangram task, cf. Krauss & Weinheimer, 1966). This preserves many aspects of face-to-face conversation, which is arguably the primary site of language use (Clark, 1996). However, because of the relatively unrestricted nature of the task, analyses of linguistic and other behaviors are performed at a coarse temporal grain, and factors such as the linguistic form of referring expressions cannot be controlled. As a consequence, the time course of effects of common ground in conversation, and its interaction with fine-grained aspects of other contributory linguistic processes, is poorly understood. This step is crucial, however, in order to understand how common ground is used during everyday real-time language comprehension.

In real-time comprehension, listeners make provisional interpretations at multiple levels of representation (e.g., lexical, syntactic, semantic, discourse model) as the linguistic input unfolds (see Gibson & Pearlmuter, 1998 and Tanenhaus & Trueswell, 1995, for reviews). Words and phrases are often open to several interpretations until the entire utterance is encountered. Comprehenders must therefore access, construct, and coordinate representations from a variety of subsystems in order to handle temporary ambiguity while still making rapid interpretation tractable. It is not clear what would constitute optimal use of common ground under these conditions. One possibility is that conversational partners continuously update their mental representations of common ground, such that a dynamically changing representation of common ground defines the domain within which utterances are typically processed. Updating common ground at this temporal grain would require participants to continuously monitor each other's utterances and actions in order to gather evidence about each other's beliefs. Alternatively, common ground might be interrogated at a coarser temporal grain. If so, many aspects of linguistic processing, including reference resolution, might take place without

initial appeal to common ground. Thus, while common ground might control language performance at a macro-level, the moment-by-moment processes necessary for production and comprehension could take place relatively egocentrically for speakers and listeners. Keysar and colleagues' monitoring and adjustment (Horton & Keysar, 1996) and perspective adjustment (Keysar, Barr, Balin, & Brauner, 2000) models adopt this approach.

In the perspective adjustment model (Keysar et al., 2000), the initial interpretation of utterances is egocentric. Common ground is used as a second-stage filter to rule out inappropriate interpretations. A number of theoretical arguments can be marshaled in support of the common ground as filter approach. Computing common ground by building, maintaining, and updating a model of a conversational partner's beliefs could be inefficient because it is extremely memory intensive. In addition, many conversational situations are constrained enough so that an individual participant's perspective provides a sufficient approximation of true common ground. Moreover, information about another's beliefs can be uncertain at best. Thus, adjustment from monitoring and explicit feedback might be the most efficient mechanism for correcting the occasional confusions or misunderstandings that arise from adopting an egocentric perspective.

Perhaps the most striking evidence that common ground might not initially restrict referential domains comes from a study by Keysar, Barr, Balin, and Brauner (1996) and Keysar et al. (2000) (see also Keysar, Barr, & Horton, 1998). Keysar et al. (2000) monitored eye movements during a referential communication task. Participants were seated on opposite sides of a vertical grid of squares, some of which contained objects. One of the participants (a confederate speaker) played the role of director and instructed the matcher addressee, who wore a head-mounted eye-tracker, to reorganize the objects in different locations in the grid. Most of the objects were in common ground on the basis of physical co-presence since they were visible from both sides of the display, but a few objects in the grid were hidden from the director's view, and thus were in the matcher's *privileged* ground. On critical trials, the director used a referring expression that referred to a target object in common ground, but that could also refer to a hidden object in privileged ground. For example, one display had three blocks in a vertical row, and the block at the very bottom was hidden. If the director said *Put the bottom block below the apple*, the underlined definite noun phrase referred to the middle block from the common perspective, but to the hidden object from the matcher's egocentric perspective. The results showed that matchers were not only just as likely to look at the hidden object as the target object, but in fact initially preferred to look at the hidden object, and on some trials even picked it up and began to move it.

Keysar et al. (1996, 1998) took this pattern of results as evidence that addressees first pursue interpretations based on an egocentric perspective, a perspective which in this task included both the hidden and visible objects. He further argued that these interpretations are monitored and then adjusted in a later stage to take into account common ground only when necessary, such as when the hidden object was selected by mistake. However, as Keysar et al. (2000) later acknowledged, these results do not demonstrate that common ground is completely ignored in initial processing.

In the critical conditions, the hidden object was always a better perceptual match for the referring expression than the visible object. For example, when the director said *Put the bottom block...*, the hidden block was the one on the absolute bottom of the display. Similarly, when the director said *Put the small candle...*, the hidden candle was the smallest candle in the display, and the visible candles were medium and large sized. In the control conditions, the hidden objects were replaced with a completely unrelated item. The data show that matchers were slower to look at the target object (e.g., the block in the middle and the medium sized candle) in the control conditions than they were to look at the hidden object (e.g., the block at the very bottom and the smallest candle) in the critical conditions. This indicates that the critical hidden objects were the most typical referents when they were present, and this typicality could have contributed to the inability of the matchers to ignore their privileged information in favor of the information in common ground.

The work reported here explored whether information provided by common ground influences the early moments of reference resolution. Experiment 1 used the head-mounted eye-tracking methodology (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) to monitor eye movements in a referential communication task in which objects were either in common ground or were privileged. In order to assess the interference of privileged information on comprehension while avoiding a confound with typicality, the potential referents were always identical. In addition, we used a different basis for common ground. The Keysar et al. (2000) experiment relied on physical co-presence between the director and the matcher as the basis for common ground. While physical co-presence is among the strongest kinds of evidence for common ground (Clark, 1992), it is somewhat unusual for objects immediately in front of two people to differ in perceptual accessibility. Furthermore, one of the most typical characteristics of common ground is that it is incrementally accumulated and adjusted interactively during conversational interaction (Clark, 1992, 1996). While the objects in the Keysar et al. study were co-present, they were not interactively entered into common ground by the conversational participants. Experiment 1 used an explicit grounding

procedure with common ground established on the basis of linguistic co-presence.

Experiment 1: Common ground and domain restriction

Participants wore a head-mounted eye tracker while they played the role of addressee in a referential communication task with a confederate speaker. (For convenience, the male confederate speaker will be called **C** and referred to as he, while participant addressees will be called **A** and referred to as she.) **C**, who was hidden behind a divider, instructed **A** to manipulate colored shapes on a display board with the pretense of getting **A**'s board to match **C**'s. **C** followed a script so that the form of the referring expressions could be controlled and linked to **A**'s eye movements and actions.

On critical trials, **C** gave an instruction containing a definite noun phrase that referred to a particular shape as a target location (e.g., *Now put the blue triangle on the red one*). On these trials there were two identical shapes (among others) already in place on the board, the target shape and the competitor shape (e.g., two triangles). The target shape was always in common ground because it had been referred to and placed on the board during the course of the matching task. We manipulated whether the competitor shape was also in common ground, or whether it was in **A**'s privileged ground. When a shape was in privileged ground, it was identified only to **A**, and she placed this "secret shape" in a space on the display board before the matching task began. In addition to manipulating the ground of the competitor shape, the target and competitor shapes were either the same or a different color. See Fig. 1 for example displays.

The competitor shape was a possible referent (e.g., of *the red one*) only when it was the same color as the target shape. Thus, we predicted that there would be no competition from a different color competitor in either common or privileged ground. The critical question was whether a same color competitor would compete when it was in privileged ground as strongly as when it was in common ground. We predicted that when the same color competitor was in common ground, addressees would consider both the target and competitor shapes as possible referents, and in fact should ask for clarifying information to determine which location was intended by **C**. The crucial condition was when the same color competitor was in privileged ground. If the language processing system makes immediate use of common ground, addressees should quickly choose the target shape in common ground and there should be little if any interference from the secret shape. However, if common ground is used to filter initially egocentric interpretations, then addressees should initially consider the secret shape equally as often as the target shape in common ground.

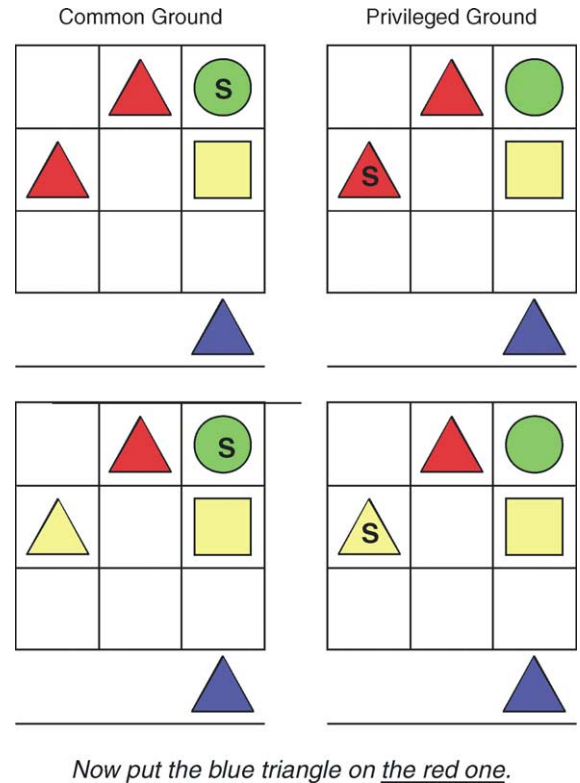


Fig. 1. Example displays and critical instruction for a single item rotated through the four conditions in Experiment 1, including when the competitor was in common ground (left column) or privileged ground (right column), and when it was the same color (top row) or a different color (bottom row). At the point of the final instruction, all of the shapes on the board were known to both the confederate speaker and the participant addressee except for the secret shape, indicated here with an **S**, which was only known to the addressee. The target shape location in this display was the topmost red triangle.

Method

Participants

Twelve undergraduates from the University of Rochester were paid for their participation. All were native speakers of English and were naive to the experimental manipulations. A trained undergraduate research assistant who was knowledgeable about the experiment played the role of the confederate.

Materials and design

A's display was a vertical wooden board with a three-by-three grid of spaces outlined in black. Space numbers (one through nine) were written in black marker in the center of each space. The board also had a resource area, which was a shelf located below the grid where the shapes needed for each trial were placed. **C**'s display board had the same three-by-three grid, but was

horizontal and made of posterboard. The objects were circles, triangles, and squares cut out of red, yellow, green, and blue posterboard.

The task was for C to set up his board according to written instructions, and then tell A how to arrange the shapes on her board so that it matched his own. Most of the shapes were in common ground and known to both interlocutors, but one shape was privileged information known only to A. This shape was known as the “secret shape.” This shape was truly a secret since C could not determine what it was while the experiment was underway.

Instructions for each trial were typed on index cards and placed in business envelopes. C’s instructions contained the configuration of shapes for his board as well as the script he would follow for that trial; A’s instructions said what her secret shape was and where to put it on the display board. A’s envelope also contained seven shapes, five of which she would need for the trial. C chose his shapes for each trial from a box located on his side of the table.

The confederate’s scripts for the critical trials consisted of the following series of instructions: (1) a question asking A what shapes she had in her envelope; (2) a statement telling A which shapes she needed for that trial; (3) three separate statements instructing A to place a shape on the board in a particular location; (4) a critical instruction telling A to stack the final shape on one of the shapes that was already on the board; and (5) a statement that C was finished or, 25% of the time, a question asking A to describe where her shapes were to double-check that the boards matched. On trials where the target and competitor shapes were both in common ground and were the same color, A typically asked which location was meant. C then apologized for his mistake and told A the specific location. Note that these were the only trials where C knew what condition was being presented from a cursory inspection of his own set-up. C was trained before the study began to sound as natural as possible so that participants would not suspect he was reading a script.

Critical instructions were always of the form *Now put the [stacking shape] on the [target shape]*, and they contained a definite descriptive noun phrase with an anaphor (e.g., *the red one*) to refer to a target location. On critical trials, there were two identical shapes on the board: in one condition both of these shapes were in common ground (common ground condition), and in the other condition one of the shapes was in privileged ground (privileged ground condition). These two ground conditions were crossed in a 2×2 design with competitor color (same/different).

Twelve shape arrays were constructed and paired with an instruction sequence to create 12 items. Each item had five shapes associated with it: a secret shape, a target shape, a stacking shape, and two other shapes.

In the privileged ground conditions, the competitor shape was the secret shape; in the common ground conditions, the competitor shape was one of the other shapes. The competitor shape was always the same shape as the target, the stacking shape was always the same shape but a different color than the competitor and target shapes, and the other two shapes had random shape assignments and were a different color from the target and the competitor. The frequency of each color and shape was otherwise balanced. The target and competitor shapes were at least one vertical and one horizontal space apart, and the shape locations varied and were balanced across trials.

Three of the 12 items were assigned to each of the four experimental cells, which rotated using a modified latin square design to create the four versions of each item. Each participant was exposed to only one of the four lists, and therefore to only one version of any one item. The 12 experimental items were randomly distributed within fourteen filler items, with at least one distractor intervening every two experimental items. In order to counteract expectations about the kinds of instructions present in the experimental items, six fillers did not contain an instruction to stack one shape on top of another one, and another six fillers contained a stacking instruction, but it was the second instruction instead of the last one. Two additional fillers had two identical shapes in common ground, but the location was specified exactly (e.g., *Now put the blue triangle on the red one in area 6*) in order to counteract any expectations of having to ask for more information when there were two identical shapes in common ground.

Procedure

As were told that the purpose of the experiment was to investigate how people coordinate their communication while cooperating to perform a simple task. As were informed that C was a lab assistant, but they were misleadingly told that he was naïve to the purpose and details of this particular experiment. As were encouraged to feel free to talk to C, and to ask for clarification if they ever needed it. In order to support the cover story of his being uninformed and untrained, these instructions were directed to C as well. As were seated across a table from C, who was hidden from view by a vertical divider.

A’s eye movements were monitored using an Applied Sciences Laboratories E4000 eyetracker. Two cameras mounted on a light-weight helmet provided the input to the tracker. The eye camera provides an infrared image of the eye. The center of the pupil and the first Purkinje corneal reflection are tracked to determine the position of the eye relative to the head. Accuracy is better than 1 degree of arc, with virtually unrestricted head and body movements. A scene camera is aligned with the participant’s line of sight. A calibration procedure allows

software running on a PC to superimpose crosshairs showing the point of gaze on a HI-8 videotape record of the scene camera. The scene camera samples at a rate of 30 frames per second, and each frame is stamped with a time code. A microphone connected to the HI-8 VCR provided an audio record of each trial.

For each of the 26 trials, the experimenter handed the envelopes containing set-up instructions and shapes to **A** and **C**, and then moved into the control booth to monitor the eye-tracking equipment. **A** placed her secret shape in the designated area on the board, which established her privileged ground. While this happened, **C** arranged his shapes. Then **C** and **A** determined which four shapes were necessary for that trial, establishing common ground, and went through the matching task with those shapes. After **C** was done, **A** told **C** what her secret shape was and where to put it on his board.

There were four practice trials preceding the start of the experiment: one without a stacking instruction; one with no identical shapes and an unambiguous stacking instruction; one with identical shapes in common ground and an unambiguous stacking instruction; and one with identical shapes in common ground and an ambiguous stacking instruction. The experimenter asked **C** and **A** to double-check their boards after the first and final practice trials; on the final trial, if **A** had not asked which of the identical shapes was the target, the mistake was caught at this point, and she was encouraged to ask for more information if she was not sure where to place a shape. At the end of the experiment, **As** were asked whether they suspected that **C** had been reading instructions or whether they thought **C** had known what their secret shape was. **As** were then thoroughly debriefed.

Results

Most of the addressees reported in their debriefing that they thought that the confederate was practiced at participating in the task; however, none suspected that he was reading a script, and none thought that he knew what their secret shape was.

The data were analyzed from the videotape records using an editing VCR with frame-by-frame control and synchronized video and audio channels. Fixations were coded by noting which shape participants were looking at frame by frame, with at least three frames of looking required to constitute a fixation. Coding began with the first fixation after the onset of the point of disambiguation (the color term in the noun phrase describing the target location). Coding ended when **A** began her hand movement to place the stacking shape on the chosen location, except for trials in the common ground/same color condition. Coding for these trials ended at the moment **A** began to ask which location was meant. **As** always asked a question in this condition. Trials in which no eye movements preceded the question in this

condition were not analyzed; this removed 4 out of 36 trials (11.1%). In the same color privileged ground condition, trials were excluded from analysis if **A** asked for more information (indicating that she forgot which shape was her secret shape); this only happened on 1 of 36 trials (2.8%). Of 144 trials in the experiment, two were excluded from analysis because the participant was already fixating on the target at the point of disambiguation, three were excluded because the calibration became so degraded that fixations could not be reliably coded, two were removed due to confederate error in reading the shape placement instructions, and an additional three were removed due to participant errors with the trial procedure. Discarded trials were evenly distributed across conditions. In total, 10.4% of the collected data were excluded from statistical analysis.

Latency of target choice

We calculated the total time that elapsed before a target location was chosen (or more information was requested in the common ground/same color condition). Participants were numerically fastest to choose a target location when a different color competitor was in common ground (mean = 627 ms) or in privileged ground (mean = 685 ms). Participants took longer to choose a target when a same color competitor was in privileged ground (mean = 829 ms), and took longest to initiate a question when a same color competitor was in common ground (mean = 1176 ms). The main effects of both ground and color were significant; $F(1, 8) = 9.17$, $p < .05$, $MSE = 2729$ and $F(1, 8) = 34.58$, $p < .001$, $MSE = 41763$, respectively, as was the interaction, $F(1, 8) = 16.06$, $p < .01$, $MSE = 30686$. Crucially, when the competitor was the same color as the target, participants were faster to choose a target location when the competitor was in privileged ground compared to when it was in common ground; $F(1, 8) = 16.94$, $p < .01$, $MSE = 42672$.

These means are in line with the straightforward predictions that there would be no competition from a different color competitor shape, and competition from a same color competitor shape. The competition from a same color competitor in privileged ground appears to be less than that from a same color competitor in common ground. However, the longer latencies in the common ground/same color condition include the time that it took participants to initiate a request for more information, and arguably reflect additional speech planning that wasn't present in the privileged ground/same color condition. Thus, in order to determine the precise timing of the domain restriction, particularly whether it was present during the initial moments of processing, we examined the proportion of fixations over time.

Proportion of fixations over time

Figs. 2 and 3 present the proportion of fixations on the shapes in the display in each of the conditions over

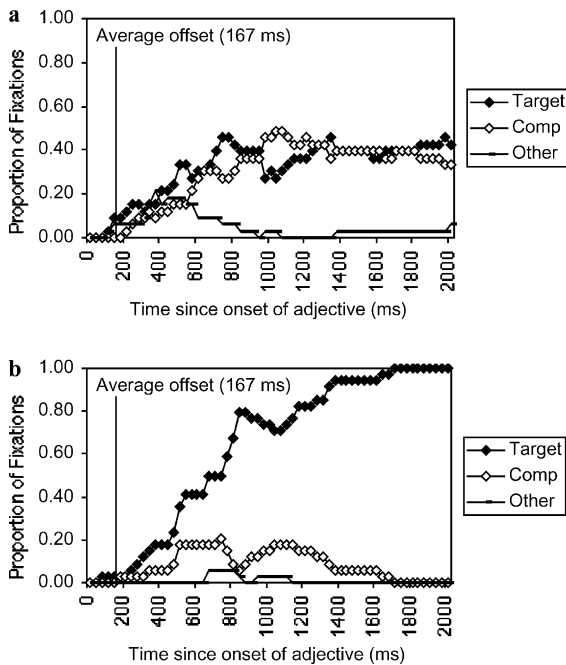


Fig. 2. Probability of fixating each object type over time in the common ground/same color competitor condition (upper graph) and the privileged ground/same color competitor condition (lower graph).

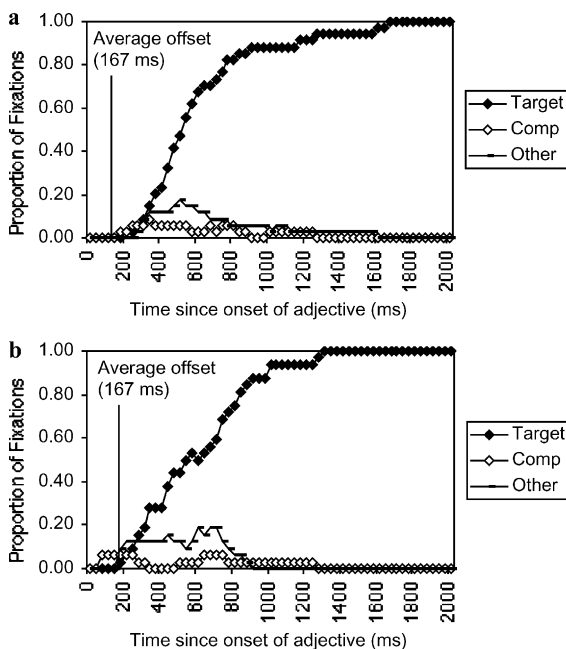


Fig. 3. Probability of fixating each object type over time in the common ground/different color competitor condition (upper graph) and the privileged ground/different color competitor condition (lower graph).

the course of 2000 ms in 33 ms intervals, beginning with the point of disambiguation. The proportions do not sum to 1.0 within a time slice because fixations to objects other than the ones in the placement area of the display were not included. At the point of disambiguation, participants were most often fixating on the stacking shape which was either in the resource area or, if they had picked it up already, in their hand. The average offset of the disambiguating region is indicated on each graph. On average the adjective in the referring expression was 167 ms in length.

When there was a same color competitor in common ground (Fig. 2a), participants initially looked roughly equally at any of the objects on the board, and within 600 ms after the onset of the disambiguating word began looking primarily at either the target or the competitor shape. Participants then often made secondary fixations on the other identical shape before asking for clarifying information. In contrast, when the competitor was a different color (Figs. 3a and b), participants more quickly fixated on the target shape, by 400 ms after the onset, and were unlikely to look at any other shape, regardless of whether the competitor was in common ground or privileged ground.

The critical question was whether a same color competitor in privileged ground would compete to the same extent as a competitor in common ground. As Fig. 2b shows, participants most often made initial looks to the target shape. Within 400 ms after the onset of the adjective, participants were fixating on the target more often than the privileged ground competitor. The proportion of fixations to the target then rose steadily and quickly. There was also a substantial number of early looks to the secret shape, more so than for another shape or a different color competitor (see Fig. 3b). The proportion of fixations to the competitor began to rise about 400 ms after the onset of the adjective, and rose along with the proportion of looks to the target until about 800 ms after the onset of the adjective. Note, however, that the proportion of looks to the privileged ground competitor was always lower than the proportion of looks to the target.

The pattern and timing of results suggest that (1) addressees used common ground as a referential constraint from the earliest moments of reference resolution and (2) there was some degree of interference from a potential referent in privileged ground. In order to evaluate these claims statistically we performed several analyses focusing on fixations within various time windows. In order to address whether addressees were statistically more likely to look at the target than the privileged ground competitor, we examined the proportion of looks to each of these objects in the time window from 200 to 800 ms. We began at 200 ms because this is the earliest point at which eye movements to an object are likely to reflect contributions from

stimulus-driven fixations. We ended at 800 ms because looks to the same color privileged ground competitor began to decrease after this point. Thus this time window allows for the most generous test of the hypothesis that there is an early egocentric stage of reference resolution. We further divided the window into two smaller 300 ms windows in order to separate the earliest looks from later looks.

Planned paired comparisons revealed that participants spent more time fixating the target shape compared to the privileged ground competitor in the 200–800 ms window; $t(1, 11) = 3.16$, $p < .01$. Moreover, participants were fixating more on the target shape in common ground in both the early 200–500 ms interval [$t(1, 11) = 2.26$, $p < .05$] and the later 500–800 ms interval [$t(1, 11) = 3.12$, $p < .01$].

Planned comparisons also revealed that participants spent more time fixating on the privileged ground competitor compared to an unrelated shape in common ground in the 200–800 ms window; $t(1, 11) = 2.38$, $p < .05$. This pattern held in both the early 200–500 ms interval [$t(1, 11) = 1.80$, $p < .05$] and the later 500–800 ms interval [$t(1, 11) = 2.39$, $p < .05$].

The amount of time that participants spent looking at the target and competitor shapes across the first 800 ms, and the shape of the proportion graphs, demonstrate that though an object in privileged ground interferes with reference resolution, it does not do so to the same extent as items in common ground. Thus, while the referential domain is not completely restricted to objects in common ground, common ground does have an immediate effect on reference resolution.

Discussion

The results of this experiment showed clear competition between the target and competitor shapes when each matched the definite referential description and each was in common ground. Under these conditions participants considered both of the shapes as possible referents and there were roughly equal looks to each of them. In contrast, when the competitor was a different color than the target and did not match the referential description, looks to the target were highest and there were no more looks to the competitor than any other shape on the board, regardless of whether the competitor was in common ground or privileged ground.

Of central interest was the condition in which the same color competitor shape was in privileged ground. Participants looked at the competitor less often than the target in common ground, but more often than an unrelated shape. In addition, participants were faster to begin looking at the target in this condition compared to when the same color competitor was in common ground.

These results provide clear evidence that common ground was not acting as the domain of interpretation in

a completely restrictive manner, as predicted by Keysar et al.'s perspective adjustment model. However, the pattern of results is inconsistent with a strong version of the model (e.g., Keysar et al., 1998) in which there is an initial stage of processing in which common ground is completely ignored. Our evidence that common ground modulated the earliest moments of reference resolution is consistent with similar studies by Arnold, Trueswell, and Lawentmann (1999) and Nadig and Sedivy (2002). Arnold et al. (1999) conducted an experiment similar to ours with a simplified display that was designed for children. They found that attentive adults experienced competition from a common ground competitor but not from a privileged ground competitor. Nadig and Sedivy (2002) used a display of four physically co-present objects, one of which was hidden from a confederate speaker. Children from 5 to 6 years of age experienced interference from a competitor in common ground but no interference from a privileged ground competitor.

These results can be naturally integrated into constraint-based accounts of language processing. Constraint-based models at a general level propose that alternative interpretations with some degree of likelihood are evaluated in parallel, based on the simultaneous and continuous integration of probabilistic evidence provided by multiple constraints. These constraints include discourse context and within-sentence structural and lexical biases, taking into account the frequencies and contingent (conditional) frequencies associated with words, categories, and structures (Jurafsky, 1996; MacDonald, 1994; Spivey-Knowlton, Trueswell, & Tanenhaus, 1993; Tanenhaus & Trueswell, 1995; Trueswell, 1996; Taraban & McClelland, 1988). For example, numerous experiments have provided evidence for the immediate role of contextual information in the resolution of a variety of within-sentence ambiguities, including syntactic attachment ambiguities and lexical thematic ambiguities (cf. Hanna, Tanenhaus, & Spivey, 2000; McRae, Spivey-Knowlton, & Tanenhaus, 1998; Spivey & Tanenhaus, 1998). However, the effects of context are graded and they interact with local constraints in predictable and quantifiable ways. When the local constraints strongly favor one interpretation, contextual constraints have weaker and/or delayed effects compared to when the local constraints are more neutral. This approach has helped integrate conflicting results from a long-standing debate about whether discourse context is the primary factor controlling initial parsing decisions or whether it is used to help evaluate and revise misanalyses after an initial stage of processing in which context is ignored.

From a constraint-based perspective, common ground can be thought of as another kind of contextual constraint that has immediate and probabilistic effects on interpretation, depending upon the strength and salience of the speaker's perspective, and its relevance

to the addressee. The constraint-based view helps reconcile the current results with the different pattern of results found by Keysar et al. (2000). In the Keysar et al. experiment, the competitor in privileged ground was a better perceptual match to the referential description than the object in common ground. Under these conditions one would expect relatively weak effects of common ground. In the current experiment, the objects in privileged and common ground were equally good referential matches. Moreover, the secret shape manipulation and the explicit grounding procedure for the shapes in common ground highlighted the importance of separating the unknown entity from the entities shared with the speaker. Thus, it is likely the participants in the current experiment were better able to keep track of the difference between the privileged and common information, and under these conditions one would expect relatively stronger effects of common ground.

Experiment 2: Common ground and remembered perspective

The research strategy introduced by Keysar and colleagues, of comparing the effects of information in common ground and privileged ground, has been quite fruitful. However, this class of experimental design is subject to a potentially problematic ambiguity of interpretation. Privileged ground objects are, by necessity, not referred to by confederates. Therefore, it is unclear whether addressees' preference for referents in common ground arises because they are taking into account the perspective of the speaker, or whether they are performing a kind of probability matching, keeping track of the likelihood with which specific referents are mentioned. This is not to say that the representation of common ground, or the mechanism by which it might exert its effects, is fundamentally distinct from probability matching, but rather that probability matching is a computation that can be done egocentrically.

The goal of Experiment 2 was to determine the time course with which an addressee can utilize information taken from a speaker's perspective when (1) it conflicts with perceptually salient information in the addressee's privileged ground and (2) conflicting perspectives are not confounded with likelihood or recency of reference.

Perspective-taking is thought of as the fundamental task of successful communication, whether a perspective includes another person's time, place, and identity, or their conceptualizations, conversational agendas, or knowledge (for a recent review of conversational perspective-taking see Schober, 1998). Moreover, it is an essential component of the construction of a joint perspective in conversation, which is required during the process of establishing common ground. Most psycho-

linguistic research into perspective-taking has focused on written and spoken language production, with methodologies that only allow coarse-grained temporal analyses (e.g., Bard et al., 2000; Brown & Dell, 1987; Clark & Wilkes-Gibbs, 1986; Garrod & Anderson, 1987; Horton & Gerrig, 2002; Lockridge & Brennan, 2002; Schober & Brennan, 2001). The experiments presented here, along with the work of Keysar and colleagues and a recent study on partner effects by Metzger and Brennan (in press), are the only fine-grained time course analyses of perspective-taking during spoken comprehension.

We manipulated the perspectives of the conversational participants in a referential communication task such that the domain of interpretation was different from each conversant's perspective; that is, from the speaker's point of view only one subset of objects would make sense as the domain in which reference would be interpreted, while for the addressee a different set of objects would potentially constitute the domain of interpretation. The participant addressee (A) was again seated across a table from a confederate speaker (C) who was completely hidden from view by a vertical divider. (In this case, the confederate speaker was female and will be referred to as she, and participant addressees will be referred to as he.) The experimenter placed four objects on a shelf on A's side of the table, and then named them in order from left to right to inform C of their identity. C repeated the object names in the same order to ground them and firmly establish her perspective, and then instructed A to pick one of them up and place it in one of two areas on the table surface (see Fig. 4).

We manipulated the precise point in the spoken instructions where the referent of the noun phrase became unambiguous with respect to the visual display. This *point of disambiguation* was varied by taking advantage of two linguistic factors: the different uniqueness conditions carried by definite and indefinite descriptions, and the contrastive property of adjectives. To give an example using a definite description, consider a display such as the top array of objects in Fig. 4 in which there are two sets of two identical objects: two jars, one with olives and one without, and two martini glasses, one with olives and one without. As the instruction *Pick up the empty martini glass* unfolds, the addressee cannot identify the referent as the martini glass without olives until relatively late in the instruction. *The empty* could refer to either the empty jar or the empty martini glass in this display, and it is not until the addressee hears *martini* that the referent is disambiguated. Now consider the same instruction in combination with a display such as the middle array of objects in Fig. 4 in which both jars are empty, and one martini glass is empty while the other martini glass has olives in it. In this case, the intended referent could be disambiguated upon hearing

Definite Conditions

Late Disambiguation/Matching Perspective Description:
*There's an empty jar, a jar with olives in it,
 and an empty martini glass,
 and a martini glass with olives in it*



Early Disambiguation/Matching Perspective and
 Mismatching Perspective Description:
*There are two empty jars,
 and an empty martini glass,
 and a martini glass with olives in it.*

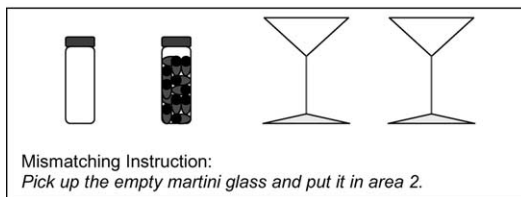
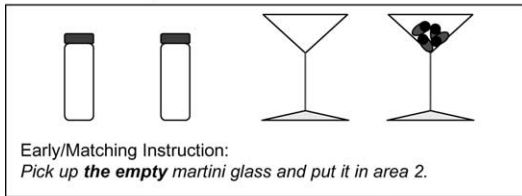


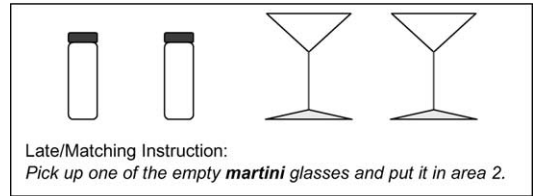
Fig. 4. Example descriptions, displays, and critical instructions for the definite late disambiguation/matching perspective, early disambiguation/matching perspective, and mismatching perspective conditions in Experiment 2. Points of disambiguation in the instructions are indicated in bold.

the empty. The definite article *the* signals a uniquely identifiable referent, which eliminates the set of jars as potential referents and locates the referent within the set of martini glasses, and *empty* uniquely identifies the martini glass without olives. Indefinite instructions such as *Pick up one of the empty martini glasses* were also combined with displays to produce late and early points of disambiguation, as shown in the top and middle arrays of Fig. 5. Studies by Eberhard, Spivey-Knowlton, Sedivy, and Tanenhaus (1995), Sedivy, Tanenhaus, Chambers, and Carlson (1998), and Chambers, Tanenhaus, Eberhard, Filip, and Carlson (2002), have all demonstrated that listeners assign reference incrementally, such that this type of point of disambiguation manipulation affects the time course of reference resolution.

We created situations where the point of disambiguation was either late or early as described above (e.g.,

Indefinite Conditions

Late Disambiguation/Matching Perspective Description:
*There are two empty jars,
 and two empty martini glasses.*



Early Disambiguation/Matching Perspective and
 Mismatching Perspective Description:
*There's an empty jar,
 a jar with olives in it,
 and two empty martini glasses.*

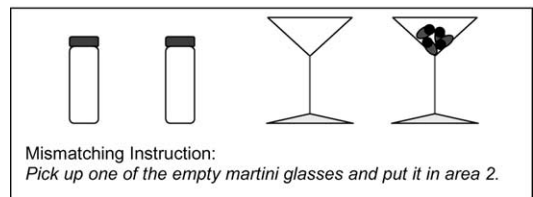
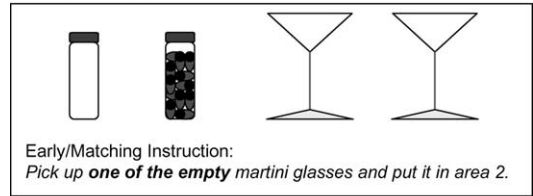


Fig. 5. Example descriptions, displays, and critical instructions for the late disambiguation/indefinite matching perspective, early disambiguation/matching perspective, and mismatching perspective conditions in Experiment 2. Points of disambiguation in the instructions are indicated in bold.

martini versus *the empty*) and where the perspectives of the participant addressee and the confederate speaker either matched or mismatched. In the matching perspective conditions, the experimenter described the objects to C accurately. In order to create circumstances where the perspectives of C and A mismatched, the experimenter described the early disambiguation displays inaccurately (from A's perspective) to C. Mismatching perspectives for both the definite and indefinite instructions were achieved by describing the objects with the modification switched between the sets. For example, for the definite instruction *Pick up the empty martini glass*, the objects were described as two empty jars, an empty martini glass, and a martini glass with olives in it, but the group really consisted of an empty jar, a jar with olives in it, and two empty martini glasses. See the bottom array of objects in Figs. 4 and 5 for the mismatching perspective conditions.

With matching perspectives, we expected to see a clear point of disambiguation effect depending on the uniqueness properties of the objects in the display. In the late disambiguation conditions, addressees should identify the target relatively slowly; early fixations should be equally distributed to both the target set of objects (e.g., the set of martini glasses) and the competitor set of objects (e.g., the set of jars), and looks to the target should not rise until after the onset of the object name. This condition is a baseline, since under all circumstances the latest point at which disambiguation can occur is at the object name; that is, there is always at least one object that matches the description of the intended referent, disregarding the definiteness of the referring expression. In the early disambiguation conditions, addressees should identify the target more quickly; fixations to the target set of objects should begin soon after the onset of the article and adjective, and there should be few looks to the competitor set of objects.

The crucial question was what would happen when there was a mismatch between C's and A's perspectives in the context of a display that would, under matching conditions, normally provide an early point of disambiguation. Taking the speaker's perspective in these conditions requires remembering what C thinks the sets of objects are like while ignoring conflicting perceptual information.¹ If use of common ground information is delayed, then addressees should initially interpret the referring expression from their own perceptual perspective, which conflicts with what the speaker believes. Thus, there should be early looks to the competitor set of objects (e.g., to the single empty jar in response to *the empty*), and a delay in the identification of the intended referent. However, if addressees are able to adopt the speaker's perspective, then reference resolution should still show an advantage compared to the late disambiguation conditions.

¹ What we are calling perspective-taking in Experiment 2 is the ability of the addressee to take into account what the speaker does and does not know about the objects in the display, similar to Experiment 1. Our paradigm is insufficient to determine whether or not addressees are actually modeling the speaker's knowledge as they do this. It is possible, for example, that addressees are adopting the strategy of pretending that the objects are different than they really are in the mismatching perspective conditions. Note that making this mental transformation requires remembering that two objects in the display are different than they really are, since addressees do not know which set of objects will contain the target or whether the referring expression will be definite or indefinite. However, the extent to which perspective-taking in circumstances like these involves this kind of pretending, and the extent to which these circumstances generalize to everyday perspective-taking, remains an open question.

Method

Participants

Eighteen undergraduate and graduate students from the University of Rochester were paid for their participation. All were native speakers of English and were naive to the experimental manipulations. The confederate was a trained undergraduate research assistant who was also naive to the experimental manipulations. A trained graduate student, also naive to the mismatching trials, was the confederate for four subjects.

Materials and design

A's display was a laminated wood shelf 36 in. in length, 8 in. in depth, and 6 in. high. The shelf was divided with dark tape into four 8 inch wide spaces, with one set of two spaces on each end and 4 in. in the middle of the shelf dividing the sets. The vertical divider between C and A had a taped cross in the center 18 in. above the surface of the shelf to provide a neutral fixation point at the start of each trial.

The objects were identical sets (pairs) of everyday household and food items, such as tissue boxes, balloons, glasses, candy bars, and CD cases. The sets could vary with respect to being empty or full, being closed or open, or being unopened or opened. For example, a set of two vases could either be empty or contain a flower, a set of two recipe boxes could either be closed or open, and a set of two packages of crackers could either be unopened or opened.

C had a written list of the object names and the critical instruction for each trial, unknown to A. In this way the object references were controlled and C did not have to rely on her memory over the course of the experiment. C never knew when the objects were being described accurately or inaccurately.

Critical instructions were always of the form *Pick up [determiner adjective object-name] and put it in area [1/2]*, and they either used the definite *the* or the indefinite *one of the*. Definite instructions referred to a single unique object within a set. Indefinite instructions referred to either object within an identical set. The adjective in the critical instructions was always *empty*, *closed*, or *unopened*. The object name was either a single head noun (e.g., *envelope*), or a compound noun or noun phrase containing other modification (e.g., *martini glass* or *deck of cards*).

Eighteen object groups were created and combined with the instructions to create 18 critical items. Each item had four objects: two sets of two identical objects, all of which could vary along the same modifying dimension (e.g., being closed/open). Within each item, one object set was the target set, and the other was the competitor set. To avoid cohort effects, the onsets of the object names in the target and competitor sets were different.

There were three kinds of displays for each item. The late disambiguation/matching perspective displays had a

point of disambiguation at the object name, and were always described accurately. Definite instructions were combined with target and competitor sets of objects that each had one unique object (e.g., one empty object and one filled object). Indefinite instructions were combined with target and competitor sets of objects that each had identical objects (e.g., all objects were empty).

In early disambiguation/matching perspective displays, the experimenter described the objects accurately, and the point of disambiguation occurred at the determiner/adjective. Definite instructions were combined with a target and competitor set of objects where only the target set had a unique object (e.g., only the target set had one empty and one filled object). Indefinite instructions were combined with a target and competitor set of objects where only the target set had identical objects (e.g., only the target set had two empty objects).

In mismatching perspective displays, the experimenter described the objects inaccurately. The modification between the target and competitor sets was switched, such that definite instructions were paired with the indefinite matching display, and indefinite instructions were paired with the definite matching display. Overall, therefore, the experiment had a 2×3 design with definiteness (definite/indefinite) and display (late disambiguation/matching perspective; early disambiguation/matching perspective; mismatching perspective) as the independent variables.

Three of the 18 items were assigned to each of six experimental cells, which rotated using a modified latin square design to create the six versions of each item. Each participant was exposed to only one of six lists and therefore to only one version of any one item. The 18 experimental items were randomly distributed within 18 filler items, with at least one distractor intervening every two experimental items.

Three types of fillers counterbalanced the experimental items. None of the filler displays were described inaccurately in order to limit the number of mismatching perspective trials to the six critical items. In order to contribute more modifiers, one group of six fillers used color as the adjectival modification, and a second group of six fillers used a big/small distinction. These 12 fillers mirrored the 18 critical conditions, having equal definite and indefinite instructions with both early and late disambiguation displays. Six more fillers, two each for the empty/full, closed/open, and unopened/opened modification, along with half of the color and big/small fillers, referred with a definite reference to the “marked” object in the target set (e.g., the equivalent of the martini glass that had olives in it). These 12 definite references to a marked object counterbalanced the nine definite references to an unmarked object (e.g., the empty martini glass) in the critical trials. In order to reduce the memory load for critical items when a mismatching perspective had to be remembered, the marked object was always on

the right within its set, and the unmarked object was always on the left. However, this only affected reference in the definite conditions, since in the indefinite conditions both objects were unmarked, and is also counterbalanced by the greater number of definite references to the marked object on the right in the filler items. Finally, over the course of the experiment, the target set was equally as often on the left and right sides of the shelf.

Procedure

When **As** arrived at the laboratory, **C** was introduced to them but then was asked to leave the room while the experiment was explained. **As** were informed that **C** was a lab assistant, and were told that she would be interacting with them to perform the task but was naive to the details of this particular experiment. **As** were told that the purpose of the experiment was to investigate how people coordinate their communication while cooperating to perform a simple task. They were then given a cover story that included a description of the kinds of feedback people give to each other while conversing, as well as the kinds of small mistakes that can be made, such as producing the wrong name for an object.

As were told that four objects would be placed in front of them and named, and that **C** would repeat back the objects to make sure she got them right and then would ask **A** to pick one of them up and move it to one of two spaces on the table surface. **As** were told that most of the time the objects would be described accurately, but that sometimes a “mistake” would be made and the experimenter would describe them a little differently. **As** were told that there were two conditions in the experiment, one where they could fix these small mistakes and one where they couldn’t, and that they were in the condition where they could not talk to **C** to fix the mistakes. They were instructed to just remember what **C** thought the objects were like on those trials so that they could perform the task as quickly and easily as possible, and they were assured that it would not be impossible for them to choose an object when mistakes were made. **As** were also told that they would fill out a questionnaire at the end of the experiment that would ask them questions about the task.

A’s eye movements were monitored in the same manner as Experiment 1, using an Applied Sciences Laboratories E5000 eyetracker. For each of the 36 trials, four objects were placed on the shelf and named from left to right by the experimenter, who then moved out of **A**’s sight into the control booth for the rest of the trial. **C** then repeated back the names of the objects, and asked **A** to pick one of them up and move it to either area 1 or 2. Thus, the entire group of objects was always described twice, from left to right, before the critical instruction, once by the experimenter and once by **C**. The two descriptions of the objects were identical, with each of the four objects described individually (e.g., *There’s an empty*

jar and a jar with olives in it, and ...) except when they were identical, in which case they were grouped with the quantifier *two* (e.g., *There's two empty jars, and ...*).

As were told that they could look at the objects during the experimenter's description and when C was repeating the names, but that they should then look up at the cross on the divider before C said which object to move. C paused on purpose after repeating the object names in order to allow time for this.

There were three practice trials preceding the start of the experiment so that A could become familiarized with the procedure, especially looking at the cross before the start of the critical instruction. The practice trials consisted of a definite early disambiguation unopened/opened trial, a definite mismatching color trial, and a definite late disambiguation big/small trial.

At the end of the experiment, As were given a questionnaire asking them to rate the difficulty of the task and whether they thought that C knew when mistakes had been made. As were then thoroughly debriefed.

Results

Five participants' data had to be replaced: two because the eye-tracking record was too unstable to code, and three because they consistently neglected to look at the cross before the critical instruction or were confused about when to look at the cross for more than a third of the experiment. Only two participants asked for more explanation during the practice mismatching perspective trial.

The participant addressees in this experiment said that they did not suspect that the confederate was reading a script, and did not think that the confederate knew when the experimenter was describing the objects inaccurately. The rating questionnaires indicated that the task was not particularly demanding. As found the task easy, even on the trials when they had to keep in mind what C thought the objects were (means < 2, where 1 was very easy and 5 was very difficult). In addition, As did not think it would have been much easier to perform the task if they had been able to talk to C (mean = 3.8, where 1 was much easier and 5 was not at all easier). As reported that they were somewhat likely to correct the mistakes if they had been allowed to (mean = 3.1, where 1 was not at all likely and 5 was very likely).²

² Research with confederates always raises the question of how the results might change if the naive participant was skeptical about either the confederate or the cover story for the experiment. Our surveys suggested that participants accepted both the confederate and the cover story. Moreover, we cannot think of any confounds that would have been introduced if participants were actually more skeptical than they let on. Crucially, the confederate in Experiment 2 was blind to the experimental manipulations on each trial. With the exception of a subset of the critical trials, as detailed in the methods section, this was also true for the confederate in Experiment 1.

Fixations were scored by noting which space on the shelf participants were fixating, beginning with the first fixation after the onset of the referring expression, and ending with the fixation prior to A initiating a reaching movement to pick up an object. Trials in which As were not fixating the cross at the onset of the critical instruction were not analyzed; this removed 15 trials (4.6%). In addition, three trials (0.9%) were excluded from analysis because of a degraded calibration, and three trials (0.9%) were removed due to a confederate production error (a repair within the critical instruction). The discarded trials were evenly distributed across conditions. In total, 6.5% of the collected data was excluded from statistical analysis.

Figs. 6 and 7 present the proportion of fixations in each condition over the course of 2000 ms in 33 ms intervals, beginning with the onset of the referring expression. (There were fixations after 2000 ms, but the patterns are visible within this time period.) The proportions do not always sum to 1.0 because any fixations to objects other than those in the placement area of the display were not included; at the beginning of the critical instruction participants were fixating on the cross. The average onset and offset of the definite or indefinite determiner, adjective, and object name are indicated on each graph with light vertical lines. In the definite conditions, the average duration of the determiner was 63 ms, the adjective was 334 ms, and the object name was 592 ms. In the indefinite conditions, the average duration of the determiner was 290 ms, the adjective was 347 ms, and the object name was 702 ms.

In the matching perspective conditions, looks to the target object should begin to separate from looks to other potential referents at a later point for the late disambiguation condition compared to the early disambiguation condition. Comparison of Figs. 6a and b and Figs. 7a and b, confirms this prediction. In the definite late disambiguation condition (Fig. 6a), fixations to either of the two objects in the target set did not begin to rise until around 900 ms after the onset of the determiner, and fixations to the single target object did not begin to diverge from all other potential referents until the end of the object name, approximately 1100 ms after the onset of the determiner. In contrast, in the definite early disambiguation condition (Fig. 6b), fixations to the target began to diverge just after the onset of the object name, about 600 ms after the onset of the determiner. Similarly, in the indefinite late disambiguation condition (Fig. 7a), participants did not begin looking more often at one of the objects in the target set until around 1000 ms after the onset of the determiner. In contrast, in the indefinite early disambiguation condition (Fig. 7b), fixations to an object in the target set began to diverge earlier at around 800 ms. Note that for the indefinite conditions, there was no identification of a single target object, since either of the identical objects in

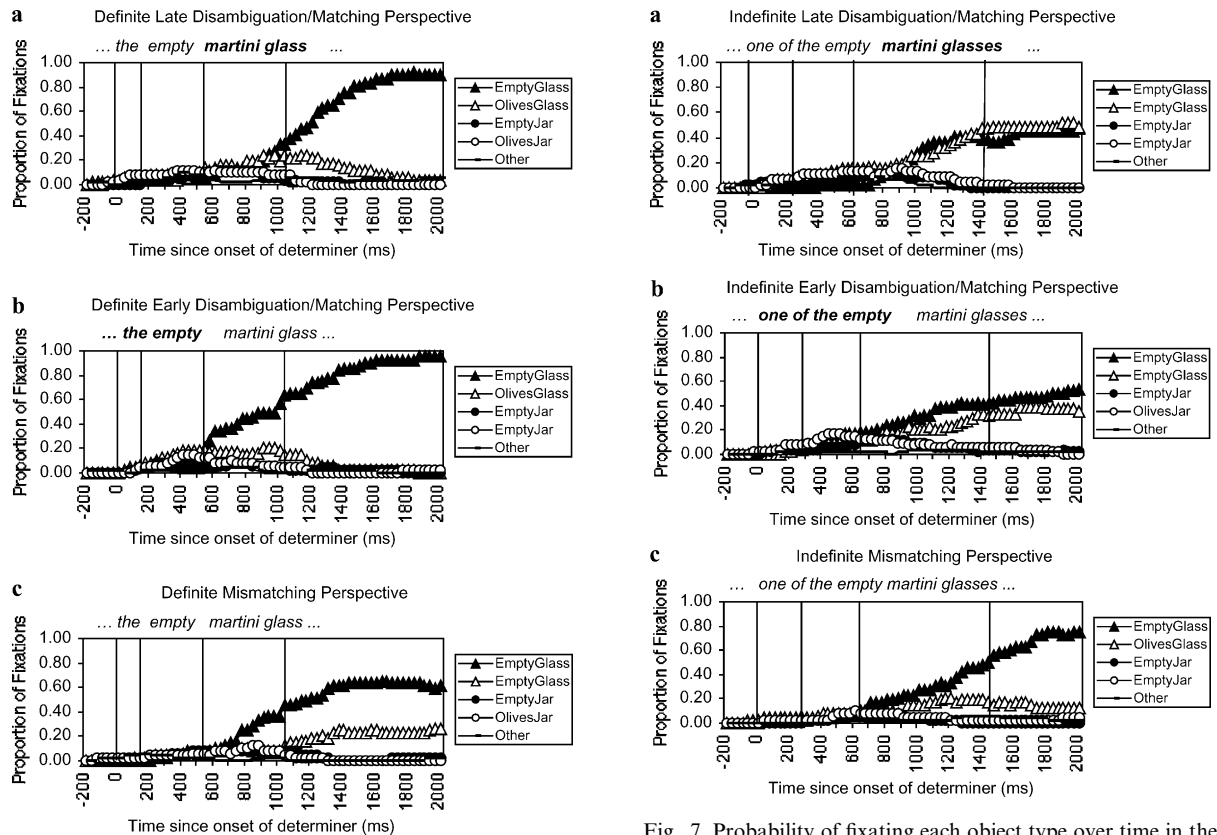


Fig. 6. Probability of fixating each object type over time in the three definite referring expression conditions. Graph (a) shows the late disambiguation/matching perspective condition, including the target (EmptyGlass), the other object in the target set (OlivesGlass), the two objects in the competitor set (EmptyJar, OlivesJar), and any other place (Other). Graph (b) shows the early disambiguation/matching perspective condition, including the target (EmptyGlass), the other object in the target set (OlivesGlass), the two objects in the competitor set (EmptyJar, OlivesJar), and any other place (Other). Graph (c) shows the mismatching perspective condition, including the two empty objects in the target set (EmptyGlass), the single empty object in the competitor set (EmptyJar), the full object in the competitor set (OlivesJar), and any other place (Other). Points of disambiguation in the instructions are indicated in bold.

the target set (e.g., either of the two empty martini glasses) was a valid choice for the indefinite referring expression, and there was a similar rise in the proportions of fixations for both target set objects. Considering that it takes about 200 ms to program and launch an eye movement it is clear that the eye movements in the late conditions were not being programmed until after the onset of the object name, while the eye movements in the early conditions were being programmed before the onset of the object name.

The crucial question is how the pattern for the mismatching perspective conditions compares to the late

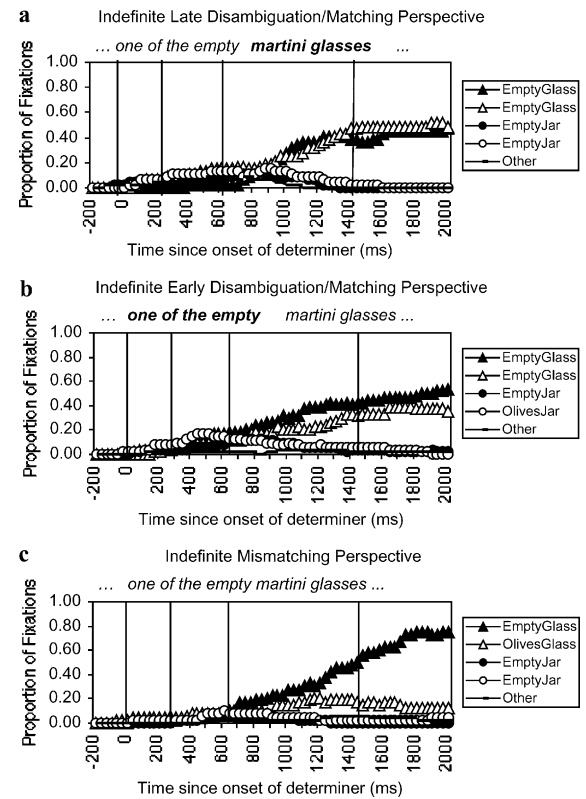


Fig. 7. Probability of fixating each object type over time in the indefinite referring expression conditions. Graph (a) shows the late disambiguation/matching perspective condition, including the target set objects (EmptyGlass), the competitor set objects (EmptyJar, OlivesJar), and any other place (Other). Graph (b) shows the early disambiguation/matching perspective condition, including the target set objects (EmptyGlass), the competitor set object on the left (EmptyJar), the competitor set object on the right (OlivesJar), and any other place (Other). Graph (c) shows the mismatching perspective condition, including the single empty object in the target set (EmptyGlass), the full object in the target set (OlivesGlass), the two full objects in the competitor set (EmptyJar, OlivesJar), and any other place (Other). Points of disambiguation in the instructions are indicated in bold.

and early matching perspective conditions. Figs. 6c and 7c show that for both the definite and indefinite mismatching conditions there were few looks to the competitor objects, and looks to the target object or objects began to diverge from the other potential referents earlier than in the late disambiguation conditions. In the definite mismatching condition, looks to either of the target objects (recall this is the condition where there are actually two empty martini glasses) began to diverge about 700 ms after the onset of the determiner, well before the 900 ms divergence point for the corresponding late condition; in the indefinite mismatching condition, looks to the single target object (there was only one empty martini glass) began to diverge about 800 ms after

the onset of the determiner, again well before the 1000 ms divergence point for the corresponding late condition.

In order to evaluate the statistical reliability of this data pattern, a variable we will call ‘windows’ was added as a factor to the design, dividing the data into ten 200 ms windows from the onset of the determiner to 2000 ms following this onset. Planned interaction comparisons were conducted within the windows on the proportion of time spent looking at the target, either of the competitor objects, and the other object in the target set. For the indefinite conditions, and the definite mismatching perspective condition, the target set contained two identically modified objects, but the object on the left was designated as the target for comparison to the definite conditions. Both independent variables, definiteness and display, were within subjects and items. Item analyses were not conducted because variations in items reflected differences in the characteristics of the display, not differences in the form of the instructions. Since display type was rotated through subject groups, we included list as a factor, but it had no systematic main effects or interactions (see Chambers et al., 2002).

Within individual windows, there were significant main effects of both display and definiteness. Although there was a small length difference between the definite and indefinite instructions (the determiner and adjective region was longer for the indefinites), this difference did not affect the eye movement patterns for each of the display types, as display did not interact with definiteness. Therefore, we will address the effects of display and definiteness separately, beginning with the display effects.

There were no significant effects of display on the proportion of target looks in window 1 (0–200 ms) or window 2 (200–400 ms), all p values $> .1$. Within window 3 (400–600 ms), there was a significant main effect of display [$F(2, 24) = 8.28$, $p < .01$, $MSE = .022$] due to a higher proportion of time spent looking at the target in the early/matching condition (mean = 17%) than either the late/matching condition (mean = 3%) [$F(1, 12) = 10.23$, $p < .01$, $MSE = .033$] or the mismatching perspective condition (mean = 7%) [$F(1, 12) = 7.56$, $p < .05$, $MSE = .025$]. This indicates that addressees were able to begin identifying the target object earliest when the point of disambiguation was at the determiner/adjective in the early/matching condition.

Within window 4 (600–800 ms), there was again a significant main effect of display [$F(2, 24) = 13.06$, $p < .001$, $MSE = .028$]. The proportion of target looks was higher in the early/matching condition (mean = 29%) than either the late/matching condition (mean = 9%) [$F(1, 12) = 17.05$, $p < .001$, $MSE = .042$] or the mismatching perspective condition (mean = 16%) [$F(1, 12) = 7.60$, $p < .05$, $MSE = .037$]. Crucially, however, the proportion of target looks was also higher for the mismatching

condition than for the late/matching condition [$F(1, 12) = 19.64$, $p < .001$, $MSE = .005$], indicating that addressees began identifying the target earlier in the mismatching perspective condition than in the late/matching condition.

It was not until window 5 (800–1000 ms) that the proportion of looks to the target in the late/matching condition began to rise. There was a main effect of display [$F(2, 24) = 7.57$, $p < .01$, $MSE = .032$], but this time although the proportion of target looks was higher for the early/matching condition (mean = 38%) than the late/matching condition (mean = 22%) [$F(1, 12) = 15.44$, $p < .01$, $MSE = .031$] and the mismatching perspective condition (mean = 28%) [$F(1, 12) = 6.42$, $p < .05$, $MSE = .028$], the latter two conditions did not differ, $p > .1$.

The pattern of target looks in window 5 was repeated within window 6 (1000–1200 ms) and window 7 (1200–1400 ms). Display had a main effect that was marginal in window 6 [$F(2, 24) = 3.20$, $p = .06$, $MSE = .054$] and significant in window 7 [$F(2, 24) = 4.83$, $p < .05$, $MSE = .100$]. The proportion of target looks in the early/matching condition (window 6 mean = 51%; window 7 mean = 60%) was higher than in the late/matching condition (window 6 mean = 39%; window 7 mean = 46%); marginal in window 6 [$F(1, 12) = 4.58$, $p = .054$, $MSE = .054$] and significant in window 7 [$F(1, 12) = 6.70$, $p = .05$, $MSE = .055$]. The proportion of target looks in the early/matching condition was also higher than in the mismatching condition (window 6 mean = 38%; window 7 mean = 37%); marginal in window 6 [$F(1, 12) = 4.33$, $p < .06$, $MSE = .063$] and significant in window 7 [$F(1, 12) = 9.07$, $p < .05$, $MSE = .105$]. The proportion of target looks in the late/matching and mismatching conditions did not differ in either window 6 or window 7, all p values $> .1$.

Finally, in window 8 (1400–1600 ms) and beyond, the effect of display was no longer significant, all p values $> .1$, reflecting the equally high proportion of target object looks in all of the display conditions at this point.

There were no significant effects of display within any of the individual windows on the proportion of time spent looking at either of the objects in the competitor set, all p values $> .1$. The proportion of competitor looks was consistently very small, ranging from 2% to 11%, and addressees did not look at the objects in the competitor set more often in the mismatching perspective condition than in the early/matching or late/matching conditions.

In sum, reference resolution occurred earlier in the early disambiguation/matching perspective condition compared to the late disambiguation/matching perspective condition, replicating previous findings by Eberhard et al. (1995) and Sedivy et al. (1998). Addressees were rapidly able to take the speaker’s perspective into account. The intended referent was identified earlier, and there were no more competitor looks, in the

mismatching perspective condition compared to the late disambiguation/matching perspective condition. However, there was some cost associated with perspective-taking; target looks began to rise later for the mismatching perspective condition compared to the early disambiguation/matching perspective condition.

As expected, there was an effect of definiteness because addressees could choose a single target object in the definite conditions, but could choose either object in the target set in the indefinite conditions. Beginning in window 4, there was a main effect of definiteness on the proportion of looks to the target object (mean definite = 24%; mean indefinite = 13%) [$F(1, 12) = 7.24$, $p < .05$, $MSE = .035$]. This reflects the slower rise in target looks in the indefinite condition, which had a longer determiner and adjective region, and the fact that there were two possible target objects in the indefinite condition drawing fixations. This effect of definiteness was significant within the remaining individual windows, all p values $< .01$. Thus, addressees looked at the target object more often in the definite conditions than in the indefinite conditions. Note that this was true even for the definite mismatching perspective condition, where both objects in the target set were identical, indicating that addressees preferred to choose the object whose initial linguistic description matched the referring expression.

Turning to the proportion of looks to the other object in the target set, the effect of definiteness did not become significant until window 6 (1000–1200 ms), where the proportion was higher in the indefinite condition (mean = 24%) than in the definite condition (mean = 17%); $F(1, 12) = 5.28$, $p < .05$, $MSE = .026$. Beginning in window 7 (1200–1400 ms), not only was the main effect of definiteness significant [$F(1, 12) = 15.50$, $p < .01$, $MSE = .037$], but the main effect of display was also significant [$F(2, 24) = 4.60$, $p < .05$, $MSE = .027$], and these two factors also interacted [$F(2, 24) = 3.95$, $p < .05$, $MSE = .051$]. This pattern of significant main effects and an interaction was also found in window 8 (1400–1600 ms), window 9 (1600–1800 ms), and window 10 (1800–2000 ms), all p values $< .001$. Addressees were in general looking at the other object in the target set more often in the indefinite conditions than in the definite conditions, but this was only true for the early/matching and late/matching conditions. The interaction of definiteness and display was due to the fact that in the definite conditions, the proportion of looks to the other object in the target set was generally higher in the mismatching perspective condition than in the early/matching [all p values $< .05$] or sometimes the late/matching condition [p marginal in windows 7–9 and $p < .05$ in window 10], while in the indefinite conditions, the proportion of looks to the other object in the target set was generally lower in the mismatching perspective condition than in either the late/matching [all p values $< .05$] or the early/matching

condition [$p < .05$ in windows 8–10]. This makes sense given that the mismatching perspective conditions paired definite instructions with indefinite displays, and indefinite instructions with definite displays. In the definite mismatching condition, there were two objects in the target set that matched the referring expression, and addressees looked at the other object in the target set relatively often; in the indefinite mismatching condition, there was only a single target object that matched the referring expression, and addressees did not look at the other object in the target set as often.

This pattern of definiteness effects gives some insight into the manner in which addressees were able to take into account the perspective of the speaker once they had restricted the domain of interpretation to the target set of objects. In the definite mismatching perspective condition, either object was an equally good perceptual match for the portion of the referring expression after the determiner. However, addressees chose the object on the left in the target set more often than the other object in the target set, making it their final choice 72% of the time. Thus, it is apparent that the object whose description matched the confederate's referring expression, for example the empty martini glass on the left that had been described as empty, was the preferred interpretation, even though there was no basis for this choice perceptually. It is also apparent, though, that perceptual information did play a role in reference resolution, since the other object in the target set was sometimes chosen. Furthermore, in the indefinite mismatching perspective condition, both objects had been described identically, but only one of the objects in the target set perceptually matched the portion of the referring expression after the determiner. There was a strong bias towards the perceptually matching object in the target set, much more so than in the definite mismatching condition, with participants choosing it 92% of the time. Interestingly, addressees did choose the other object in the target set 8% of the time, indicating that they were willing to take the speaker's perspective to the extent that they sometimes acted as if the objects were exactly as they had been described.

Discussion

Addressees were able to use the definiteness of referring expressions given by a confederate speaker to rapidly restrict the domain of interpretation. Definite referring expressions restricted addressees' attention to the set of display objects with a unique member, and indefinite expressions restricted their attention to the set of display objects with identical members. When the two sets of display objects had the same uniqueness characteristics, the point of disambiguation was at the object name, and the rise in fixations to the target set of objects did not begin until relatively late. When early

disambiguation was possible before the object name, participants rapidly fixated the correct set of objects, and then showed a quick rise in fixations to the single target object for the definite referring expressions, and an equal rise in fixations to either of the objects in the target set for the indefinite referring expressions. These results provide further evidence that the uniqueness requirements of determiners rapidly act as a constraint in the dynamic updating of domains of reference, consistent with the results of Chambers et al. (2002).

Critically, the same pattern held with mismatching perspectives. Although addressees had to remember that the speaker thought that the objects were different than they really were, they were still able to make rapid use of information provided by uniqueness and contrast. There was some cost associated with perspective taking, but the results patterned earlier than the conditions where there was a late point of disambiguation and the perspectives matched. Furthermore, there were no more looks to the competitor set of objects in this condition than in the matching perspective conditions, indicating that there was little interference from the egocentric perceptual perspective.

Interestingly, the results from both the definite and indefinite mismatching perspective conditions showed that information from two different sources, the referential grounding and the perceptual scene, had differential effects on addressees' behavior once they had identified the relevant set of objects. When there were actually two objects that matched the speaker's definite referring expression, addressees quickly identified the single object that had the matching description, but showed some interference from the other object which was a perceptual match to the reference. When there was only one object that matched the speaker's indefinite referring expression, addressees also quickly identified a single object, but this time it was the one that was also the perceptual match to the reference, showing some interference from the other object that had the matching description. This suggests that while addressees were able to remember and use the speaker's perspective immediately, their own perceptual perspective also had significant effects.

The combined effects of perceptual and linguistic match present problems for an alternative explanation based on *linguistic precedence*. One might argue that the addressees simply adopted the name for the object assigned by the experimenter. This explanation would predict that the participants' initial eye movements would be determined by the linguistic match between the confederate's request and the name of the object used by the experimenter. The linguistic precedence account makes the correct predictions for the definite mismatching perspective condition because only one object (e.g., *an empty martini glass*) was a possible linguistic precedent for a definite noun phrase, and this object was

the preferred interpretation. However, it cannot account for the data from the indefinite mismatching perspective condition because there were two linguistically acceptable precedents (e.g., *two empty martini glasses*) for an indefinite expression, and yet the object which perceptually matched the referring expression was the immediately preferred interpretation. In conjunction with the results from Experiment 1, therefore, this data pattern provides further support that multiple sources of information, including information from an egocentric perceptual perspective as well as information that comes from taking another person's linguistically co-present perspective, all play an immediate and probabilistic role during comprehension.

General discussion

The current experiments make several contributions to our understanding of how common ground is used in real-time language comprehension. Experiment 1 replicated Keysar et al.'s (2000) finding that an addressee cannot ignore perceptually salient objects in privileged ground when the objects match a speaker's referential description. However, potential referents in common ground were preferred over those in privileged ground from the earliest moments of reference resolution. The design of Experiment 2 eliminated potential confounds due to recency and/or likelihood of mention that are intrinsic to the design introduced by Keysar et al. (2000) and that we used in Experiment 1. The results demonstrated that addressees assigned reference as soon as a potential referent was uniquely identifiable given the information provided by contrast and definiteness, replicating and extending the findings of Eberhard et al. (1995), Sedivy et al. (1998), and Chambers et al. (2002). Crucially, the same pattern of results held even when using information about contrast and definiteness required the addressee to hold the perspective of the speaker in memory, and the speaker's perspective mismatched salient perceptual information from the addressee's privileged ground.

Taken together, the current results confirm Keysar et al.'s (2000) conclusion that common ground does not completely circumscribe the referential domain for referring expressions. Even in Experiment 2, where addressees were making arguably optimal use of information from common ground, there was still evidence that processing was influenced by perceptually salient information from privileged ground. However, we found no evidence for an initial stage of processing where addressees ignored information from common ground, as would be predicted by the strong version of the monitoring and perspective adjustment model.

How can we reconcile the results reported here with some of the striking demonstrations of speaker and

addressee egocentricity provided by Keysar and colleagues? We have proposed that common ground can be most fruitfully viewed as a probabilistic constraint within the framework of constraint-based processing models (e.g., MacDonald, 1994; Tanenhaus & Trueswell, 1995). In constraint-based models, different information sources or constraints each contribute probabilistic evidence for alternative interpretations during processing. The constraints are weighted according to their salience and reliability, and are integrated with each other in parallel, causing the alternative interpretations to compete with each other. Factors such as speaker perspective can be incorporated into constraint-based models through expectation-based constraints, such as the likelihood that a speaker will refer to a particular entity (cf. Arnold, 2001). This approach predicts that effects of common ground will vary with the strength and saliency of information about common ground as well as its task-specific relevance. Indeed, Keysar, Barr, and Lim (2002) recently conducted an experiment showing that participants tended to show more egocentric processing when mutual belief was not a consistent cue for reference resolution. Their revision of the monitoring and perspective adjustment model proposes that addressees adapt or satisfice, relying on common ground to the extent that it is informative. This approach is in line with our constraint-based account.

Thus far constraint-based models have been applied primarily to situations where the strength of constraints, even contextual constraints such as the prior discourse, can be estimated from experience-based factors such as frequency and plausibility. In conversational interactions in which the participants have behavioral goals, however, the state of the context must be based upon the speakers' and addressees' intentions and actions. Under these circumstances the strength and relevance of different constraints will have to be computed with respect to continuously updated contextual models because the relevancy of constraints changes moment by moment. Developing formal models of dynamically updated context will be a major challenge for constraint-based models of comprehension, as well as for other classes of models. We should note that this challenge is similar to that faced by models of perception and action as they seek to accommodate the increasing evidence that basic perceptual processes are strongly influenced by attention and intention, which are guided by behavioral goals.

We suggest that focusing on privileged ground, and on cognitive representations of common ground, may underestimate the degree to which addressees track and use speaker-based information. Our results demonstrate that addressees clearly can use limited capacity cognitive resources to consciously track a speaker's knowledge and hold onto this information in memory. However, this may not be the natural way that addressees keep track of the speaker's perspective. Factors such as eye-gaze, ges-

ture, head position, and postural orientation are likely to provide cues that allow participants to track each other's perspectives, attentional states, and intentions, without requiring memory intensive cognitive models of mutual belief. Just as basic low-level cognitive mechanisms such as priming may be at least partially responsible for many phenomena that support coordination (e.g., lexical entrainment), tracking the perspective of a conversational participant might be accomplished in part via basic low-level mechanisms which social primates use to monitor each other during interaction. Whether or not these mechanisms result in fully-developed cognitive models of the intentions and beliefs of a conversational partner is an important question; but, this question is orthogonal to the more basic question of how conversational partners achieve the coordination necessary for successful real-time communication.

References

- Arnold, J. E. (2001). The effect of thematic roles on pronoun use and frequency of reference continuation. *Discourse Processes, 31*, 137–162.
- Arnold, J.E., Trueswell, J.C., & Lawentmann, S.M. (1999, November). Using common ground to resolve referential ambiguity. Paper presented at the 40th Annual Meeting of the Psychonomic Society. Los Angeles, CA.
- Bard, E. A., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language, 42*, 1–22.
- Barwise, J. (1989). *The situation in logic*. Stanford, CA: Center for the Study of Language and Information.
- Barwise, J., & Perry, J. (1983). *Situations and attitudes*. Cambridge, MA: MIT Press.
- Brennan, S.E. (1990) Seeking and providing evidence for mutual understanding. Unpublished doctoral dissertation, Stanford University.
- Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 22*, 482–493.
- Brown, P. M., & Dell, G. S. (1987). Adapting production to comprehension: The explicit mention of instruments. *Cognitive Psychology, 19*, 441–472.
- Chambers, C. G., Tanenhaus, M. K., Eberhard, K., Filip, H., & Carlson, G. (2002). Circumscribing referential domains in real time language comprehension. *Journal of Memory and Language, 47*, 30–49.
- Clark, E. V., & Clark, H. H. (1979). When nouns surface as verbs. *Language, 55*, 430–477.
- Clark, H. H. (1992). *Arenas of language use*. Chicago: University of Chicago Press.
- Clark, H. H. (1996). *Using language*. New York: Cambridge University Press.
- Clark, H. H. (1997). Dogmas of understanding. *Discourse Processes, 23*, 567–598.
- Clark, H. H., & Brennan, S. E. (1989). Grounding in communication. In L. Resnick, J. Levine, & S. Teasley (Eds.),

- Perspectives on socially shared cognition*. Washington, DC: American Psychological Association.
- Clark, H. H., & Haviland, S. (1977). Comprehension and the given-new contract. In R. Freedle (Ed.), *Explaining linguistic phenomena*. Washington: Hemisphere.
- Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. H. Joshe, B. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10–63). Cambridge, UK: Cambridge University Press.
- Clark, H. H., & Shaefer, E. F. (1987). Collaborating on contributions to conversations. *Language and Cognitive Processes*, 2, 19–41.
- Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22, 1–39.
- Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, 24, 409–436.
- Epstein, R. (1998). Reference and definite referring expressions. *Pragmatics & Cognition*, 6, 189–207.
- Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, 27, 181–218.
- Gibson, E., & Pearlmutter, N. J. (1998). Constraints on sentence comprehension. *Trends in Cognitive Science*, 2, 262–268.
- Hanna, J. E., Tanenhaus, M. K., & Spivey, M. J. (2000). Integrating contextual and sentential constraints in ambiguity resolution. Manuscript submitted for publication.
- Horton, W. S., & Gerrig, R. J. (2002). Speakers' experiences and audience design: Knowing when and knowing how to adjust utterances to addressees. *Journal of Memory and Language*, 47, 589–606.
- Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, 59, 91–117.
- Jurafsky, D. (1996). A probabilistic model of lexical and syntactic disambiguation. *Cognitive Science*, 20, 137–194.
- Keysar, B., Barr, D.J., Balin, J.A., & Brauner, J.S. (1996, November). Common ground: An error correction mechanism in comprehension. Paper presented at the 37th Annual Meeting of the Psychonomic Society. Chicago, IL.
- Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, 11, 32–37.
- Keysar, B., Barr, D. J., & Horton, W. S. (1998). The egocentric basis of language use: Insights from a processing approach. *Current Directions in Psychological Science*, 7, 46–50.
- Keysar, B., Barr, D.J., & Lim, S. (2002, November). Satisficing in language processing: The adaptive theory of language use. Paper presented at the 43rd Annual Meeting of the Psychonomic Society. Kansas City, MO.
- Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology*, 4, 343–346.
- Lockridge, C. B., & Brennan, S. E. (2002). Addressees' needs influence speaker's early syntactic choices. *Psychonomic Bulletin and Review*, 9, 550–557.
- MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes*, 9, 157–201.
- McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, 38, 283–312.
- Metzing, C. A., Brennan, S. E. (in press). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language*.
- Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science*, 13, 329–336.
- Schober, M. F. (1998). Different kinds of conversational perspective-taking. In S. R. Fussell, & R. J. Kreuz (Eds.), *Social and cognitive psychological approaches to interpersonal communication*. Mahwah, NJ: Lawrence Erlbaum.
- Schober, M.F., Brennan, S.E. (in press). Processes of interactive spoken discourse: The role of the partner. In A.C. Graesser, M.A. Gernsbacher, S.R. Goldman (Eds.), *Handbook of discourse processes*. Hillsdale, NJ: Lawrence Erlbaum.
- Searle, J. R. (1969). *Speech acts. An essay in the philosophy of language*. Cambridge: Cambridge University Press.
- Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1998). Achieving incremental processing through contextual representation: Evidence from the processing of adjectives. *Cognition*, 71, 109–147.
- Smith, N. V. (Ed.). (1982). *Mutual Knowledge*. London: Academic Press.
- Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning Memory and Cognition*, 24, 1521–1543.
- Spivey-Knowlton, M. J., Trueswell, J. C., & Tanenhaus, M. K. (1993). Context effects in syntactic ambiguity resolution: Discourse and semantic influences in parsing reduced relative clauses. *Canadian Journal of Experimental Psychology*, 37, 276–309 [Special issue].
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. E. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 632–634.
- Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension. In J. Miller, & P. Eimas (Eds.), *Handbook of cognition and perception*. San Diego, CA: Academic Press.
- Taraban, R., & McClelland, J. L. (1988). Constituent attachment and thematic role assignment in sentence processing: Influences of context-based expectations. *Journal of Memory and Language*, 27, 597–632.
- Trueswell, J. C. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language*, 35, 566–585.